User Authentication from Web Browsing Behavior

Myriam Abramson
Naval Research Laboratory, Code 5584
Washington, DC 20375
myriam.abramson@nrl.navy.mil

David W. Aha
Naval Research Laboratory, Code 5514
Washington, DC 20375
david.aha@nrl.navy.mil

Abstract

As anticipated in *True Names* by Vernor Vinge, identity has been recognized as our most valued possession in cyberspace. Attribution is a key concept in enabling trusted identities and deterring malicious activities. As more people use the Web to communicate, work, and otherwise have fun, is it possible to uniquely identify someone based on their Web browsing behavior or to differentiate between two persons based solely on their Web browsing histories? Based on a user study, this paper provides some insights into these questions. We describe characteristic features of Web browsing behavior and present our algorithm and analysis of an ensemble learning approach leveraging from those features for user authentication.

1 Introduction

The problem of user identity is one of the fundamental and still largely unresolved problems of cyberspace, testing the boundary between trust and privacy. Multiple approaches have been proposed to solve this problem through consolidated password schemes (e.g., OpenID (Thibeau and Reed 2009), Firefox's Persona (Mills 2011)). On the other hand, the popularity of social media such as Facebook and Twitter have made possible the availability of large amount of spontaneous online usage behavior ripe for analysis and individual search history patterns are already used by Google to personalize search results. Reality mining (Pentland and Pentland 2008) captures unconscious patterns of behavior through signals obtained from wearable mobile computing devices to reveal personal characteristics in order to shape human interaction. As our interaction with the Web becomes more natural and even mediates our interaction with others (Turkle 2012), we claim that Web browsing behavior can be rich enough to uniquely characterize who we are through unconscious behavioral patterns and authenticate ourselves with a cognitive personal fingerprint.

Attribution is broadly defined as the assignment of an effect to a cause. We differentiate between authentication and identification as two techniques for the attribution of identity. Authentication is defined as the verification of claimed identification (Jain, Bolle, and Pankanti 1999). Identification involves recognition as a one-to-many matching problem while authentication is a one-to-one matching problem. While biometric methodologies strive to provide instant authentication results, this paper focuses on the continuous authentication problem where authentication is made over time through the monitoring of activities.

The paper is organized as follows. In Section 2, we briefly describe prior research on the modeling of Web browsing behavior and attribution in cyberspace. In Section 3, we present our descriptive analysis of the different features of Web browsing behavior from clickstream data obtained through a user study. In Section 4, we present our empirical analysis on authenticating users with classifiers trained from individual features and introduce our algorithm for an ensemble of classifiers trained from subsets of those features. Our conclusions and future work suggestions are in Section 5.

2 Related Work

Marketers have long been interested in understanding Web interaction behavior (Atterer, Wnuk, and Schmidt 2006) in order to design Web sites that entice visitors to finish their Web session with a checkout of their shopping cart. *Behavioral targeting* is an approach used by advertisers (e.g., DoubleClick) that tracks Web behavior to deliver advertisements which match an individual's semantic profile defined by content-related preferences and interests. Research in this area has concentrated on identifying the demographic characteristics of a behavior such as age, gender, and income rather than authenticating a single individual (Goel, Hofman, and Sirer 2012). There has also been some research on understanding online browsing behavior from an aggregate perspective in order to identify influential websites in user navigation patterns (Kumar and Tomkins 2010).

In contrast to semantic patterns, syntactic patterns characterize Web browsing based strictly on session and navigation features. They include the burstiness of pageviews, the number of page revisits, and the number of pages between revisits (Kumar and Tomkins 2010). As an illustration of burstiness, it was noted in (Kumar and Tomkins 2010) how the inter-arrival time cumulative distribution between any visit to a particular URL and the previous visit fits a logarithmic function across users. On the server side, visitors are independent of each other so the distribution of visits can follow a Poisson distribution. It is not clear if this type of distribution also fits the time distribution on the user side since the webpages visited are not independent from each other. We

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
User Authentication from Web Browsing Behavior	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Code 5584,4555 Overlook Ave., SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
2013 Florida Artificial Intelligence Research Society Conference, St. Pete Beach, FL, 22-24 May.

14. ABSTRACT
As anticipated in True Names by Vernor Vinge, identity has been recognized as our most valued possession in cyberspace. Attribution is a key concept in enabling trusted identities and deterring malicious activities. As more people use theWeb to communicate, work, and otherwise have fun, is it possible to uniquely identify someone based on their Web browsing behavior or to differentiate between two persons based solely on their Web browsing histories? Based on a user study, this paper provides some insights into these questions. We describe characteristic features of Web browsing behavior and present our algorithm and analysis of an ensemble learning approach leveraging from those features for user authentication.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**6**	

will show another illustration of burstiness on the user side. In addition, the length of a session (both time and number of pages visited) and the starting time and day of the week also characterize user syntactic patterns.

The attribution problem in cyberspace has been addressed in several ways mainly by leveraging from features in the browser (e.g., history stealing, cookies) or accessing datasets containing partially identifying information. For example, de-anonymization in social networking websites has been accomplished by computing the intersection of users from group memberships in a social network using information from hyperlinks in the browser history and knowledge about those groups (Wondracek et al. 2010). In general, unique identification is possible by cross-referencing independent information sets containing partial information with a universal set in a manner equivalent to a database join (also known as "linkage attacks"). For example, it has been possible to link medical records to individuals in voter registration records (Sweeney 1996). Some success has been reported with the classification of global syntactic features of a Web session (e.g. length of session, average time on a page) per user (Padmanabhan and Yang 2006) aggregated over several sessions. It has also been shown that authorship of content can be determined from stylometric features on an internet scale threatening anonymity (Narayanan et al. 2012) but this type of attribution depends on published content. Research in predicting user behavior in cyberspace has also been focused on improving tasks such as information retrieval (Armstrong et al. 1995). For example, based on the content of the current webpage and a user's original search keywords, the most relevant hyperlinks in the page are highlighted to guide selection of the next page to visit. This type of prediction is oriented toward the information presented in context to the user rather than the specific activity that a user might pursue (e.g. send an email, read a paper, etc.). In contrast to previous approaches, we address the attribution problem by leveraging both from syntactic patterns in Web browsing history and the semantic content of this history with the genre of the page.

The authentication problem has been addressed in the context of masquerade detection in computer security by modeling user command line sequences. In the masquerade detection problem, the task is to positively identify masqueraders but not to positively identify a particular user. Recent experiments modeling user-issued OS commands as bag-of-words without timing information have obtained a 72.7% true positive rate and a 6.3% false positive rate (Salem and Stolfo 2010) on a set of 15000 commands for 70 users grouped in sets of 100 commands. In that work, a one-class support vector machine (SVM) (Schölkopf et al. 2000) was shown to produce better performance results than threshold-based comparison with a distance metric. We extend the results of this work to individual feature sets of Web browsing behavior and in combination with an ensemble.

3 Web Browsing Modeling

Logging of spontaneous clickstream data in our user study consisted of recording through custom-built browser extensions (Firefox and Chrome) the timestamp and the URL that was visible at the time by the user (i.e., *pageview*). The data was parsed offline to minimize interference with the user. Ten subjects (2 females and 8 males) participated in this study during the course of their work for one month. For clarity, we only show the results of the same 3 users in our figures. The population was fairly homogeneous and rated themselves highly "Web savvy." The following features, which are detailed later, were extracted from the data: day-of-week, time-of-day, pauses (below 5 mins), burstiness (below 10 mins), time between revisits, and genres (i.e. page types). The number of pageviews per user varied from 1200 to 12000. Web browsing behavioral data is noisy and requires some pre-processing for analysis. Noise occurs due to distortion from the network behavior, errors in accessing URLs, and automatic page insertion in the browser. Future work will mitigate those problems.

The clickstream data is parsed into "sessions" where a session is defined as a continuous stream of pageviews delimited by pauses greater than 30 minutes as in (Kumar and Tomkins 2010). The number of sessions for our users varied from 42 to 205. The length of a session averaged from 14 to 131 pageviews. User sessions are the data points in our study of Web behavior. We distinguish between global session features and internal session features as explained below.

3.1 Global Session Features

Standard global session features capture characteristics of a session across pageviews. They include day-of-week (DOW) and time-of-day (TOD) distributions. Since the advent of teleworking and flex time, these features are not uniform across workers. Figure 1 illustrates three users and their patterns of weekly online activity aggregated for all sessions. User 3 is the only one not active during the weekend. Figure 2 shows for the same three users their patterns of hourly online activity aggregated across all sessions. User 2 is mostly active in the morning while User 1 is active after dinner.

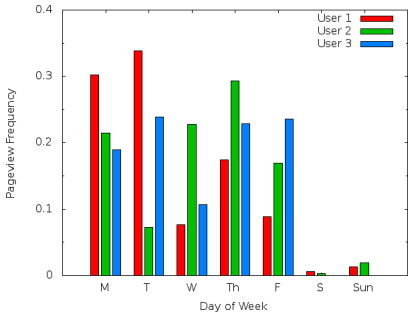

Figure 1: Daily activity patterns for three users aggregated across all sessions

Other global session features in our empirical study include the total number of pageviews, the average duration of pageview, and the number of unique pageviews.

3.2 Internal Session Features

An internal session feature captures characteristics of pageviews within a session.

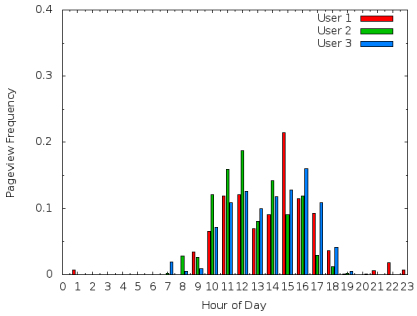

Figure 2: Hourly activity patterns for three users aggregated across all sessions

Pauses Pauses are the time spent by the user on a webpage. It is computed as the difference between the timestamp of two consecutive pageviews. Like other human activities, pause profiles follow the power law distribution (Barabasi 2005). Consequently, we can fit this data with an exponential function. Figure 3 shows the exponential fit of pause profiles below 5 minutes for three users. This data fit function can be used to obtain the probability of the next pageview and act as a signature by which to compare pause distributions. Differences between users are more pronounced for shorter pauses.

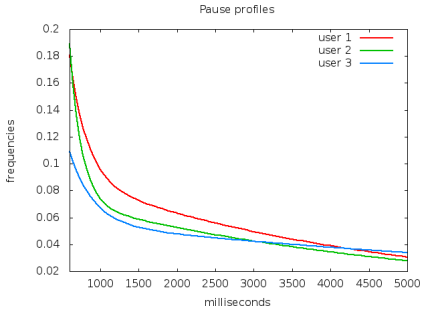

Figure 3: Pause profiles below 5 mins for three users aggregated across all sessions truncated to the first 5 seconds

Burstiness Burstiness, as a characteristic of human behavior, follows the power law distribution. In (Barabasi 2005) burstiness is explained as a consequence of our decision process in prioritizing tasks. It is computed as the change in pause time between pageviews or second order pause time (Kwok 2012). While burstiness patterns are fairly uniform across users for longer pause changes, they can be quite different for shorter pause changes as illustrated in Figure 4.

Time between revisits How often is a webpage revisited? Some webpages were found to play a role similar to stop words in a sentence (Montgomery and Faloutsos 2001). The rate at which webpages are revisited may also serve as an indicator of user identity. The revisit rate averaged between 28% to 46% among our users. Figure 5 illustrates the time between revisits (under 6 mins) profile for three users. There are large differences mainly in the shorter intervals of time.

Genres Encoding is necessary to obtain reusable patterns of behavior. We encode the semantic and stylistic content

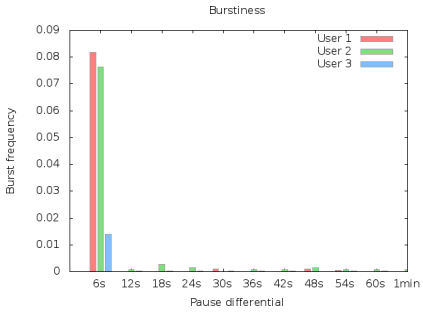

Figure 4: Burstiness profile below 1 min aggregated across all sessions for three users

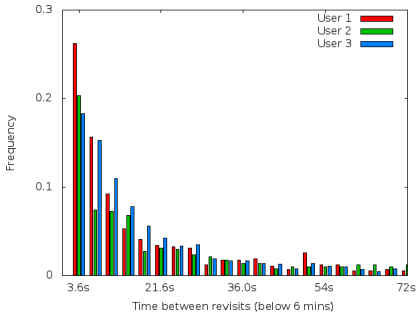

Figure 5: Time between revisits (under 6 min.) profile for three users across all sessions truncated to the first 72s

of webpages into genres. Genres are functional categories of information presentation. In other words, genres are a mixture of style, form, and content. For example, books have many genres such as mystery, science-fiction, fiction, and biography. Similarly, webpages have evolved their own genres (e.g, blog, homepage, article). Basically, the genre of a document is tied to its purpose and reflects social conventions for disseminating and searching information. We claim that genres are more indicative than topics for distinguishing Web browsing behavior. For example, some people are more frequent visitors of discussion forums (e.g. reddit) than blogs (e.g. wordpress) regardless of content. However, genres and topics do combine in important ways (e.g. spam is a combination of content and style).

We used the Diffbot page classifier [1] to classify pages into genres. Diffbot is a web service that currently categorizes webpages into 21 pages. There are several problems in using a third party web service especially one that is in beta mode. Although we expect that the quality of the categorization will improve as Diffbot matures, the main problems are certificate errors (some of which could be resolved internally by loading the certificates or via automatic trust configuration), external errors (which include errors that a user could have experienced), errors due to the Web service itself (10% of all accesses), the limitation in the number of requests per month, and control over the page types. Figure 6 illustrates the genre profiles for three users. There are large differences between users in the genre of pages visited. No strong linear correlation was found between genres and pauses so we can't infer the time spent on a webpage from its genre.

[1] http://www.diffbot.com

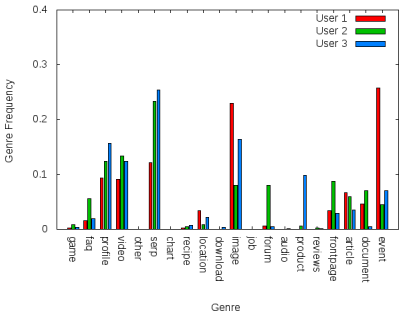

Figure 6: Genre profiles for three users (excluding errors) aggregated across all sessions

4 Empirical Study

The goal of this study is to verify the claim that users can be authenticated from their Web browsing behavior. All experiments were conducted in the Weka machine learning workbench (Hall et al. 2009) augmented by our own ensemble algorithms. We extracted the features of Web browsing behavior described above from each user session and aggregated them into one feature vector. A user's dataset consisted of all sessions collected for that user. For each user, we compared the false rejection rate (FRR) (i.e., false negative rate) and the false acceptance rate (FAR) (i.e., false positive rate) for classifiers derived from each feature set and an ensemble classifier composed of classifiers based on a weighted random sample of those features. FRR results were obtained using cross-validation on the user's dataset while FAR results were obtained by applying the classifier obtained on a dataset containing the data of all the other users. Note that FRR results will be better in practice.

4.1 One-Class Classification

One-class classification is pertinent in the context of classification with only positive examples where negative examples are hard to come by or do not fit into a unique category. Some applications for one-class classification include anomaly detection, fraud detection, outlier detection, authorship verification and document classification where categories are learned individually. The goal of one-class classification is to detect all classes that differ from the target class without knowing them in advance. One-class classification is similar to unsupervised learning but tries to solve a discriminative problem (i.e., self or not self) rather than a generative problem as in clustering algorithms or density estimation. Several algorithms have been modified to perform one-class classification. We used a one-class SVM available with LibSVM (Schölkopf et al. 2000) as part of the Weka machine learning toolbench. SVMs are large-margin classifiers that map feature vectors to a higher dimensional space using kernels based on similarity metrics. The optimization objective in SVMs is to find a linear separating hyperplane with maximum margin between class boundaries. In the case of a Gaussian kernel, a non-linear separating hyperplane is found that separates the class boundaries. A kernel transforms the feature space using a similarity measure to "support" vectors (i.e., instances close to decision boundaries) maximizing the margin. Formally, let x and x' be two feature vectors and Φ a feature mapping function to a higher-dimensional space, a kernel function k is defined as $k(x, x') = \Phi(x)^T \Phi(x')$. Since the number of features and number of examples (sessions) for each user is relatively small, we use the radial basis function kernel (Hsu et al. 2003) based on a Gaussian transform of the feature space with default parameters. The one-class SVM in the LibSVM library simply finds a separating hyperplane with respect to the origin as a support vector in the complement class.

Table 1 shows the results of one-class SVM classification for each user and for each feature set. The global features consists of the DOW distribution, the TOD distribution, the number of pageviews, the number of unique pageviews, and the average duration of each pageview in the session. For each session, pauses (below 5 mins), bursts (below 10 mins), and time between revisits were discretized into 100 bins. All feature distributions (DOW, TOD, pauses, bursts, revisits, and genres) were normalized. In addition, Each feature was scaled in the [-1,1] range in the training dataset (i.e., the user's dataset). FRR results are obtained with 10-fold cross-validation averaged over 10 runs while FAR results are obtained by applying the classifier trained on the entire user dataset to the data of the other users applying the feature scaling obtained during training (Hsu et al. 2003). Please note that FRR results should be better in practice.

Figure 7 aggregates the results of Table 1. It illustrates the tug-of-war between FRR and FAR outcomes and the difficulty of obtaining good results for authentication metrics. An increase in FRR is usually accompanied by a decrease in FAR and vice versa. Genres and global features were found to be good at differentiating Web browsing behavior (as evidenced by lower FAR rates) while pauses, bursts, and revisits were found to have better recognition rates (as evidenced by lower FRR rates). However, none of the individual features are good enough in isolation to authenticate a user.

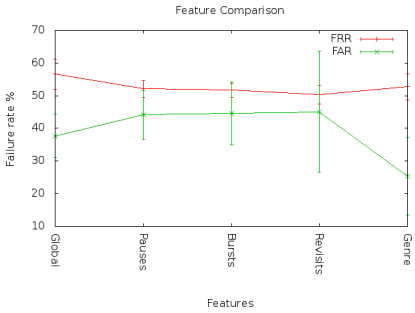

Figure 7: Average feature set results comparison

4.2 Ensemble Learning

Can we leverage collectively from those features to improve performance? Accuracy and diversity in individual classifiers were found to be necessary and sufficient conditions for high-performing ensemble of classifiers (Dietterich 2000). Furthermore, it was shown that ensemble learning does not follow Occam's razor principle stating that increased complexity decreases generalization accuracy (Ho 1998). Ensemble learning varies the type of learner or the type of input (e.g., the set of instances or features) to achieve diversity. For example, bagging (Breiman 1996) varies the set of in-

		Global Features		Pauses		Bursts		Revisits		Genres	
	#sessions	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR	FAR
User 1	98	50±1.07	35±0.0	50±0.78	30±0.0	52±1.40	37±0.0	47±0.00	32±0.00	50±1.88	28±0.0
User 2	72	52±1.63	43±0.0	51±0.94	52±0.0	50±0.97	48±0.0	48±0.94	59±0.00	48±4.43	34±0.0
User 3	86	55±1.71	33±0.0	52±0.97	37±0.0	52±0.67	41±0.0	54±1.13	50±0.00	62±6.18	22±0.0
User 4	121	56±1.37	37±0.0	51±1.25	42±0.0	51±1.16	37±0.0	50±0.92	56±0.0	49±2.79	1±00
User 5	88	59±1.59	31±0.0	50±0.87	45±0.0	50±0.70	29±0.0	45±0.00	0±0.0	55±2.26	25±0.0
User 6	181	53±1.07	33±0.0	50±0.53	51±0.0	50±0.53	52±0.0	51±0.00	60±0.0	51±0.97	23±0.0
User 7	205	55±0.92	39±0.0	51±0.57	56±0.0	51±0.52	64±0.0	51±0.84	50±0.0	51±0.79	33±0.0
User 8	42	64±2.78	44±0.0	58±1.35	41±0.0	53±1.26	45±0.0	53±1.26	48±0.0	55±3.27	20±0.0
User 9	59	60±1.41	49±0.0	55±0.97	44±0.0	51±1.66	45±0.0	53±2.27	36±0.0	54±2.51	47±0.0
User 10	44	62±2.40	27±0.0	53±2.13	43±0.0	57±0.0	48±0.0	51±1.03	60±0.0	52±3.02	20±0.0
Avg	99	56.5	37.1	52.1	44.1	51.7	44.6	50.3	45.1	52.7	25.5

Table 1: FRR and FAR results (given as percentages) obtained with a one-class SVM classifier for each feature set for each user. FRR results are averaged over 10 runs.

stances, while the random subspace method varies the set of features (Ho 1998). We use the random subspace method to vary the input features of one-class SVMs. Two-fold cross-validation on the training set evaluates the weight of a classifier used to combine the decisions of the different classifiers (i.e., self or not self) in a weighted vote. Choosing accurate classifiers is problematic here since it is easy to overfit in the one-class classification problem as a classifier choosing a class (self) at random could achieve perfect accuracy! To overcome this problem and address the diversity issue, we select a subset of the classifiers with weighted sampling. We train a fixed number of classifiers (300) each with a random subset of features (5) as a pool of classifiers to choose from. A fixed number of classifiers (107) were then selected from this pool for our ensemble. These parameters, number of classifiers, number of features and pool size, were selected empirically for good performance on User 1 without adjustment for the other users. Future work will select a variable number of features. Pauses and time-between-revisit distributions were truncated to the first 20 bins to prevent spurious features due to sparsity in the data. Algorithm 1 describes our methodology. Table 2 compares the random subspace method with a mixture of experts ensemble where the decision of the classifiers trained on the individual feature sets are combined into a weighted vote using a similar methodology.

	Mixture of Experts		Random Subspace	
	FRR	FAR	FRR	FAR
User 1	50±0.66	30±0.0	18±2.20	7±6.70
User 2	56±1.83	48±0.0	33±7.14	7±9.20
User 3	54±2.27	32±0.0	41±4.05	10±8.84
User 4	53±1.87	35±0.0	28±3.56	5±5.64
User 5	49±1.35	16±0.0	22±3.26	19±8.84
User 6	50±1.18	43±0.0	35±4.28	11±8.13
User 7	53±0.53	50±0.0	32±4.11	26±10.47
User 8	55±0.84	37±0.0	44±8.99	13±11.27
User 9	50±1.08	39±0.0	27±6.11	15±10.05
User 10	55±2.37	37±0.0	32±6.61	15±7.68
Avg	52.5	36.7	31.2	12.8

Table 2: FRR and FAR results obtained by ensemble methods of one-class SVM classifiers with weighted vote scheme.

Algorithm 1 Random subspace ensemble learning training methodology where $instances$ are the training instances, P is the pool size, cl the classifier algorithm, f the number of features, and n the ensemble size ($n \leq P$).

```
BUILDCLASSIFIER(instances, cl, P, f, n)
 features ← instances.features
 FOR i = 0 to P
  FOREACH feature in features
   feature.weight ← random
  END
  fsubset ← weight_sampling (features, f)
  // Transform instances
  fInsts ← filter (instances, fsubset)
  // Two-fold cross-validation
  eval ← cross-validate(cl,2,fInsts)
  model ← train-classifier (cl,fInsts)
  model.weight ← eval
  model[i] ← model
 END
 models ← weight_sampling (model, n)
RETURN models
END
```

The random subspace method further increases the bias of the classifiers by restricting the amount of features which in turn reduces overfitting, a major source of classification errors. There is a clear linear relationship between FRR results and the ensemble size (i.e., the number of selected learners from the pool) (Fig. 8). FAR results depend both on the ensemble size and the pool size (Fig. 9). There is a significant performance difference ($p < 0.05$) between FAR results from our random subspace ensemble learning method and from the mixture of experts method except for User 8 which recorded the least number of sessions. There is a significant difference in FRR results between the two methods for half the users, which suggests that some adjustments in the parameters for specific users might be required.

5 Conclusion

Authentication is important in scaling up the attribution of Web behavior to large number of users. Our experiments have shown that although the individual features of Web browsing behavior are not individually or collectively strong enough to authenticate and distinguish users, our random

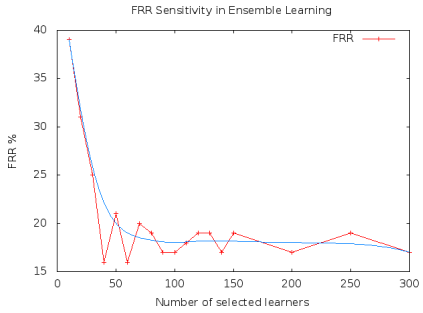

Figure 8: Sensitivity between FRR results and ensemble size with pool size of 300 and 5 features for User 1

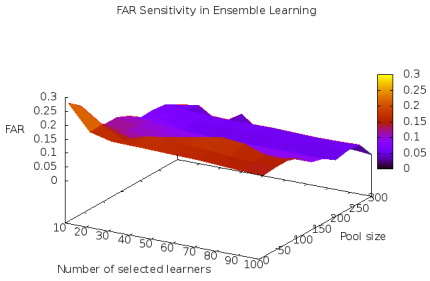

Figure 9: Sensitivity between FAR results, ensemble size and pool size in random subspace ensemble with 5 features for User 1

subspace method for ensemble learning can dramatically improve those results. Future work will include additional features as well as the exploration of additional one-class learners. Other research issues include extending our methodology to group profiles.

Acknowledgement

The authors want to acknowledge Kevin Kwok whose knowledge of the Web was critical to this project during his internship.

References

Armstrong, R.; Freitag, D.; Joachims, T.; and Mitchell, T. 1995. Webwatcher: A learning apprentice for the world wide web. In *AAAI Spring Symposium on Information Gathering from Heterogeneous, distributed environments*.

Atterer, R.; Wnuk, M.; and Schmidt, A. 2006. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the International World Wide Web Conference WWW06*, 203–212. ACM.

Barabasi, A. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211.

Breiman, L. 1996. Bagging predictors. *Machine learning* 24(2):123–140.

Dietterich, T. 2000. Ensemble methods in machine learning. *Multiple classifier systems* 1–15.

Goel, S.; Hofman, J.; and Sirer, M. 2012. Who does what on the web: Studying web browsing behavior at scale. In *Proceedings of the 26th conference on artificial Intelligence AAAI*.

Hall, M.; Frank, E.; Holmes, G.; and Bernhard Pfahringer, Peter Reutemann, I. H. W. 2009. The WEKA data mining software: an update. In *SIGKDD Explorations*, volume 11.

Ho, T. 1998. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20(8):832–844.

Hsu, C.; Chang, C.; Lin, C.; et al. 2003. A practical guide to support vector classification.

Jain, A.; Bolle, R.; and Pankanti, S. 1999. *Biometrics: personal identification in networked society*. kluwer academic publishers.

Kumar, R., and Tomkins, A. 2010. A characterization of online browsing behavior. In *Proceedings of the 19th international conference on World wide web*, WWW '10, 561–570. New York, NY, USA: ACM.

Kwok, K. 2012. User identification and characterization from web browsing behavior. Technical report, US Naval Research laboratory.

Mills, D. 2011. Introducing BrowserID: a better way to sign in. Retrieved from http://identity.mozilla.com/post/7616727542/introducing-browserid-a-better-way-to-sign-in.

Montgomery, A., and Faloutsos, C. 2001. Identifying web browsing trends and patterns. *Computer* 34(7):94–95.

Narayanan, A.; Paskov, H.; Gong, N.; Bethencourt, J.; Stefanov, E.; Shin, E.; and Song, D. 2012. On the feasibility of internet-scale author identification. In *Proceedings of the 33rd Conference on IEEE Symposium on Security and Privacy*.

Padmanabhan, B., and Yang, C. 2006. Clickprints on the web: Are there signatures in web browsing data? Technical report, Wharton School, University of Pennsylvania.

Pentland, A., and Pentland, S. 2008. *Honest signals: how they shape our world*. The MIT Press.

Salem, M., and Stolfo, S. 2010. Detecting masqueraders: A comparison of one-class bag-of-words user behavior modeling techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 1(1):3–13.

Schölkopf, B.; Williamson, R.; Smola, A.; Shawe-Taylor, J.; and Platt, J. 2000. Support vector method for novelty detection. *Advances in neural information processing systems* 12(3):582–588.

Sweeney, L. 1996. Replacing personally-identifying information in medical records, the scrub system. In *Proceedings of the AMIA Annual Fall Symposium*, 333. American Medical Informatics Association.

Thibeau, D., and Reed, D. 2009. Open trust frameworks for open government. Retrieved from `http://openid.net/government/Open_Trust_Frameworks_for_Govts.pdf`.

Turkle, S. 2012. *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Wondracek, G.; Holz, T.; Kirda, E.; and Kruegel, C. 2010. A practical attack to de-anonymize social network users. In *IEEE Security and Privacy*.